AP-2 09/475704

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : C07H 15/12, C12N 9/48, 1/20 C12N 15/00

A1

(11) International Publication Number: WO 90/00556

(43) International Publication Date: 25 January 1990 (25.01.90)

(21) International Application Number: PCT/US89/02996

(22) International Filing Date: 13 July 1989 (13.07.89)

(30) Priority data: 218,304 13 July 1988 (13.07.88) US

(71) Applicant: THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, U.S. DEPARTMENT OF COMMERCE [US/US]; 5285 Port Royal Road, Springfield, VA 22161 (US).

(72) Inventors: LOUIS, John, M. ; 2527 Ross Road ,102, Silver Spring, MD 20910 (US). OROSZLAN, Stephen ; 11411 Duryea Drive, Potomac, MD 20854 (US). MORA, Peter, T. ; 4006 West Underwood Street, Chevy Chase, MD 20015 (US).

(74) Agents: STERN, Marvin, R. et al.; Fleit, Jacobson, Cohn, Price, Holman & Stern, The Jenifer Building, 400 Seventh Street, N.W., Washington, DC 20004 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SYNTHETIC HIV PROTEASE GENE AND METHOD FOR ITS EXPRESSION

(57) Abstract

The invention is a synthetic DNA sequence for encoding a specific enzyme or protease. The protease is essential for the completion (replication) of an infective human immunodeficiency virus (HIV). The invented gene is desirable for the expression of the protease by recombinant methodology in prokaryotic and/or eukaryotic cells and the production of a commercially desirable amount of the protease for biochemical and physical characterization, necessary to find effective inhibitor of the protease, and thereby to block the production of infectious human immunodeficiency virus (HIVs).

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Fasso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CM	Cameroon
DE	Germany, Federal Republic of
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
HU	Hungary
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
RO	Romania
SD	Sudan
SE	Sweden
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
US	United States of America

SYNTHETIC HIV PROTEASE GENE AND METHOD FOR ITS EXPRESSION

BACKGROUND OF THE INVENTION

Field of the Invention

This invention relates to synthetic genes and their expression products. Specifically, this invention relates to a synthetic protease gene and its expression product.

Description of the Related Art

The presence of protease protein in purified virion preparation was shown only by immunological techniques. The HIV protease sequence together with the gag and pol sequence or fusion proteins have been expressed from viral DNA in bacteria. Examples of such disclosures include: 1. Henderson, et al., 1988, "Human Retroviruses, Cancer and AIDS: Approaches to Prevention and Therapy", D. Bolognesi Ed. Published by Alan R. Liss Inc., New York, NY. pp.135-147; 2. Debouck, et al., 1987, P.N.A.S., 84:8903-8906, and 3. Mous, et al., 1988, J. Virol, 62:1433-1436.

The primary sequences of the HIV protease has been determined by protein analysis and by the nucleotide sequence of the proviral DNA. It was thus determined that the protease is a 99 amino acid long protein encoded by a 297bp long stretch of the HIV provirus. All previous experiments on the protease gene and on its expression were carried out by utilizing nucleotide sequences cloned out from the cDNA of the provirus. The inventors' work using synthetic DNA proves that the nucleotide sequence of the provirus DNA and also the deduced aminoacid sequence are correct.

The complete nucleotide sequence of the HIV-1

proviral DNA was published by Ratner et al., 1985, Nature, 313:277-284. The sequence coding for the protease in the pol open reading frame of HIV was determined by previous analysis and corresponds to nucleotide 1609 to 1906 The N terminus and the C terminal amino-acids are proline and phenylalanine respectively. This sequence coding for the HIV-I 99 aminoacid protease is 297bp long as follows.

```
      10         20         30         40         50
CCTCAGATCA CTCTTTGGCA ACGACCCCTC GTCACAATAA AGATAGGGGG
GGAGTCTAGT GAGAAACCGT TGCTGGGGAG CAGTGTTATT TCTATCCCCC
      60         70         80         90        100
GCAACTAAAG GAAGCTCTAT TAGATACAGG AGCAGATGAT ACAGTATTAG
CGTTGATTTC CTTCGAGATA ATCTATGTCC TCGTCTACTA TGTCATAATC
     110        120        130        140        150
AAGAAATGAG TTTGCCAGGA AGATGGAAAC CAAAAATGAT AGGGGGAATT
TTCTTTACTC AAACGGTCCT TCTACCTTTG GTTTTTACTA TCCCCCTTAA
     160        170        180        190        200
GGAGGTTTTA TCAAAGTAAG ACAGTATGAT CAGATACTCA TAGAAATCTG
CCTCCAAAAT AGTTTCATTC TGTCATACTA GTCTATGAGT ATCTTTAGAC
     210        220        230        240        250
TGGACATAAA GCTATAGGTA CAGTATTAGT AGGACCTACA CCTGTCAACA
ACCTGTATTT CGATATCCAT GTCATAATCA TCCTGGATGT GGACAGTTGT
     260        270        280        290
TAATTGGAAG AAATCTGTTG ACTCAGATTG GTTGCACTTT AAATTTT
ATTAACCTTC TTTAGACAAC TGAGTCTAAC CAACGTGAAA TTTAAAA
```

The industry is lacking a synthetic DNA sequence that encodes a specific enzyme or protease which is essential for the completion replication) of an infective human immunodeficiency virus (HIV). This DNA sequence is desirable to express this protease by recombinant methodology in bacteria and or in

eukaryotic cells, and to produce enough protease for biochemical and physical characterization in order to design and produce potent inhibitors of this enzyme, and thereby to block the production of infective HIV particles.

BRIEF DESCRIPTION OF THE INVENTION

The invention is a gene for encoding a protease of human immunodeficiency virus. The gene consists essentially of a synthetic nucleotide sequence for a protease essential to infectivity of human immunodeficiency virus.

The protease is desirably a protease of HIV-1 or HIV-2 that is essential for the infectivity of these viruses.

The preferred embodiment of this inventions is a synthetic gene and the coding sequence for expression of the HIV-1 protease is represented above by the top rows of nucleotide sequence.

BRIEF DESCRIPTION OF THE DRAWING

Figure 1 presents the expressed HIV protease as analyzed in Western blot.

Figure 2 illustrates a strategy for the synthesis of the HIV-1 protease gene. The 3' overhangs are in lower case. The complementary strands (not shown) were provided with 3' overhangs to match the coding strands.

Figure 3 illustrates the induction of the gene at various periods of time.

Figure 4 illustrates the activity of the expressed protease using a synthetic peptide asa substrate.

DESCRIPTION OF THE PREFERRED EMBODIMENT

The invention is a synthetic DNA sequence for

encoding a specific enzyme or protease. The protease is essential for the infectivity of the human immunodeficiency virus (HIV). The invented gene is desirable for the expression of the protease by recombinant methodology in bacteria and or in eukaryotic cells and the production of a commercially desirable amount of the protease for biochemical and physical characterization. This characterization is necessary for the design and production of potent inhibitors of this enzyme. The invention also includes synthesis and expression of the protease gene of other retroviruses such as HIV-2, the human leukemia viruses such as HTLV I, II, and other human and animal RNA containing viruses causing leukemia sarcoma and other malignencies.

The nucleotide sequence for the preferred embodiment of this invention was obtained from a published paper by Ratner, et al., supra. The sequence in the pol open reading frame coding for the protease of HIV-1 corresponds to nucleotide 1609 to 1906. The N-terminal and the C-terminal amino-acids are proline and phenylalanine respectively. This sequence coding for the 99 aminoacid protease is 297bp long as shown above. Minor substitutions of one or more bases in this and other genes useful in this invention can produce a variant gene capable of expressing the desired protease.

This sequence was synthesized as five fragments using the DNA synthesizer. Complementary strands corresponding to these five fragments were also synthesized. The 3' overhangs of four bases were provided for appropriate sequences to efficiently ligate each of the five fragments and to provide the correct coding sequence of the protease gene.

Nucleotide ATG were added to the fragment corresponding to the 5' end of the gene and TAA at the 3' end.

A procaryotic expression vector was used to clone and then to express the synthetic sequence coding for the protease. The expression can be in prokaryotes (bacteria) or in other appropriate expression systems. Recombinant clones screened by colony hybridization using a labelled fragment (62bp) spanning the internal region of the protease gene. Positive clones were further analyzed for the size of the insert. Clones which answered positive were induced for expression and analyzed in Western blots to determine the protein product using specific antibodies. Figure 1 gives an example.

Of the clones screened so far, 3 clones have been identified to express a product of 11.5kd, reacting against specific antibodies as illustrated in Figure 1.

Conditions for the induction of a protease gene were studied in E. coli and optimized. The inventors have shown that the gene product has specific protease activity, as it is capable of cleaving both synthetic and natural substrates. The enzyme has been purified by specific column chromatographic techniques, including affinity chromotography. The method of this invention can produce enough active protease to study the structure of the protease, its mechanisms of action, with a goal of devising specific inhibitors to this enzyme, of a therapeutic application for the treatment of the diseases, such as AIDS, caused by the viruses. Other embodiments of this invention can utilize a gene to express another protease such as the following gene for the HIV-2 protease.

CCTCAATTCTCTCTTTGGAAAAGACCAGTAGTCACAGCATACATTGAGGGTCAGCCA

GTAGAAGTCTTGTTAGACACAGGGGCTGACGACTCAATAGTAGCAGGAATAGAGTTA
GGGAACAATTATAGCCCAAAAATAGTAGGGGGAATAGGGGGATTCATAAATACCAAG
GAATATAAAAATGTAGAAATAGAAGTTCTAAATAAAAAGGTACGGGCCACCATAATG
ACAGGCGACACCCAATCAACATTTTTGGCAGAAATATTCTGACAGCCTTAGGCATGT
CATTAAATCTAC

Figure 1 demonstrates the expression of the HIV protease in E. coli. Cells transformed with the synthetic sequence of HIV protease in an appropriate expression vector were induced and the bacterial lysate was electrophoresed in SDS-PAGE. After transfer of proteins into a nitrocellulose membrane, immunoblotting procedure was performed using the specific antibody to the HIV protease. Detection of Ag-Ab complex was made using I^{125} protein A. The autoradiograph lane A represents E. coli transformed with the plasmid, and lanes B and C E. coli transformed with the plasmid bearing synthetic DNA encoding the HIV protease. On the left are protein molecular weight markers in kilodalton. The 11.5 kd band is the protease.

The synthetic DNA of the invention also obviates any need to manipulate (infectious) viral material and overcomes limitations in the quantities which can be obtained by other means.

EXAMPLES

The following materials and methods were used to perform the examples.

PLASMID, BACTERIAL STRAINS, AND CHEMICALS:

Plasmid PKK233-2, a procaryotic expression vector was purchased from Pharmacia. PKK233-2 was used to transform in a laq-q host, E. coli cell JM105 or RB791. The cells were selected in M9 minimal media containing lug/ml thiamine, prior to using them for transformation. All chemicals utilized in the synthesis of oligonucleotides were from Applied

Biosystems Inc. T4 polynucleotide kinase, DNA ligase, and Klenow fragment of E. coli DNA polymerase I were obtained from New England Biolabs. Restriction endonucleases, PMSF and IPTG were from Boehringer Mannheim, Bethesda Research Laboratories and Promega respectively.

DNA SYNTHESIS, PLASMID CONSTRUCTION AND SCREENING:

DNA fragments were synthesized using a ABI DNA synthesizer (model 381A). All synthetic fragments were purified by electrophoresis in a 12% polyacrylamide/8M urea sequencing gel. DNA was visualized by UV-shadowing and full-length fragments were eluted from the gel as known in the art. The full-length fragments were checked for their purity using standard techniques.

Appropriate complementary fragments were mixed in equimolar concentrations, annealed, kinased and ligated as described elsewhere. The efficiency of ligation was monitored by polyacrylamide gel lectrophoresis. The linearized plasmid and the protease gene in appropriate concentrations were ligated and used for transformation of E. coli, JM105. Recombinant clones were screened by colony hybridization using a 62 bp fragment labelled by kinasing. Small scale isolation of plasmid DNA from the recombinant clones was performed by the boiling method and the size of the inserts was visualized by autoradiography after labelling the 3' recessed terminal using the Klenow fragment of E. coli DNA polymerase.

ANTIBODIES TO THE HIV PROTEASE

The polyclonal antibodies were raised in rabbits against (i) a complete synthetic sequence of 1 to 99 aminoacids of the HIV-1 protease and (ii) a

tridecapeptide corresponding to the C-terminus of the protease.

ANALYSIS OF THE EXPRESSED PROTEINS

E. coli cells bearing the appropriate plasmid construct were grown to log phase, induced, and lysed by sonication. Total cell extracts were analysed by $NaDodSO_4$/PAGE and subjected to immunoblot analysis.

ASSAY FOR THE ACTIVITY OF THE EXPRESSED PROTEASE:

Oligopeptides were synthesized in a Peptide Synthesizer (Applied Biosystems Model 430A), according to the method previously published (Copeland and Oroszlan, 1981). The cleavage products were analysed by RP-HPLC on a uBondapak C_{18} column (Waters Associates). Peak fractions were analysed for amino-acid composition using a Pico-Tag amino acid analyser (Waters Associates).

EXAMPLE 1

This example represents the preferred embodiment.

RESULTS:

SYNTHESIS OF THE FULL-LENGTH PROTEASE GENE:

The nucleotide sequence of the protease gene was taken from Ratner et al. The sequence in the pol open reading frame for the protease gene starts at nucleotide 1609 and ends at 1906, for coding 99 aminoacids. This sequence and its complement were synthesized as five individual fragments of approximately 60 bases as shown in Figure 2. The 3' overhangs of 4 bases (shown in lower case) were provided for the fragments to selectively ligate the appropriate fragments to form the correct coding sequence. Translational initiation codon ATG and termination codon TAA were provided at the appropriate ends of the protease gene. A sequence was added to

provide a protrusion at the 5' end of the gene, having a cohesive end compatible to the restriction enzyme site Ncol. The 5' protrustion at the 3' end of the gene was added to provide a Hind3 compatible end. The complementary strands (not shown) were provided with 3' overhangs to match the coding strands.

EXPRESSION OF THE SYNTHETIC HIV-1 PROTEASE GENE IN E. COLI

Three clones (PR-C, PR-H, and PR-J) bearing the correct coding sequence of 297bp in the expression vector PKK233-2 were analyzed for expression to select conditions for the optimal induction of the gene. Figure 3 shows examples of Western blot analysis of the gene product.

Figure 3 illustrates expression of the synthetic protease gene in E. coli. Clone PR-C bearing the coding sequence to the protease was induced for expression. The proteins (75ug of bacterial extract) were electrophoresed in a $NaDodSO_4$/PAGE transferred to nitrocellulose and subjected to immunoblot analysis using a mixture of the two protease specific rabbit polyclonal antibodies raised against (i) a complete synthetic sequence of 1-99 amino acids of the HIV-1 protease and (ii) a tridecapeptide corresponding to the C terminus of the protease. Figure 3A shows the induction of the gene with 0.4mM IPTG at various periods of time. Figure 3B shows the induction for 30 minutes. With increasing concentrations of inducer IPTG. 1-5 represent mM concentration of IPTG at 0.28, 0.56, 1.12, 2.24 and 4.48 respectively. Figure 3C shows the analysis after 60 minutes of induction with 1mM IPTG and lysing the cells in various buffers. B1 denotes lysis of cells in 50mM Tris-HCl at pH 7.0, 150mM NaCl, 1mM EDTA, 1mM PMST, 1mM DTT and 0.5 percent

NP-40. B2 is the same as B1, but without NaCl and EDTA. B3 is in 50mM potassium phosphate at pH 6.0, 1mM PMSF and 1mM DTT. B4 is the same as B3 with a pH of 6.5. Positions of protein molecular weight markers are inducated on the left in kilodaltons.

E. coli cells bearing plasmid PR-C were grown in Luria broth to an optical density of 0.4 A600nm, and then induced at various periods of time for expression from the trc promotor by adding IPTG (isopropyl-beta-D-thiog-alactopyranoside) at a concentration of 0.4mM as seen in Figure 3A. The cloned gene expressed a single, unfused protein band of 11.5kd. Expression was maximal after 30 minutes of induction. This level decreased to about 25 percent at 60 minutes. There was no detectable expression after 120 minutes of induction and at 0 minutes. This pattern of induction was similar in the other clones (PR-H and PR-J) that were analyzed (not shown).

The results of the induction for 30 minutes with varying concentrations of inducer are shown in Figure 3B. Induction with IPTG in the range of 1mM to 4mM resulted in maximum amount of expression. Similar data were obtained on clones PR-H and PR-J (not shown).

In order to select the conditions that efficiently solubilize the protease for enzymatic analysis, different buffer systems were used for the lysis of cells (clone PR-C) after optimal induction with 1mM IPTG. It was observed that sonication in a buffer system of 50mM Tris-cl at pH 7.5, 1mM DTT, 1mM PMSF and 0.5% nonidet P-40 released 50 to 70 percent of the protease in the soluble fraction (Figure 3C). This was estimated by Western blot analysis aliquots of soluble extract and insoluble pellet for the content of the expressed product.

DEMONSTRATION OF SPECIFIC PROTEOLYTIC ACTIVITY

Figure 4 illustrates the activity of the expressed protease using a synthetic peptide as a substrate. Protease assays were carried out with 22.5ug of bacterial lysate at $37^{\circ}C$ obtained from clone PR-C, induced (A,B,C), uninduced (D), and control cells bearing just the plasmid PKK233-2 (data not shown). The nonapeptide was used as a substrate in reaction buffer (0.25 M potassium phosphate), pH 7.0, 0.5 percent (v/v) NP 40, 5 percent (v/v) glycerol, 5 mM Dithiotreit and 2 M NaCl. Aliquots of 25 ul each were taken at 0 hours (A), 1 hour (B) 3 hours (C) and 6 hours (D) analyzed by RP-HPLC. S denotes the substrate and P1 and P2, cleavage products 1 and 2 respectively.

To assess the activity of the cloned HIV-1 Protease a synthetic nonapeptide corresponding to the HIV-1 p17-p24 cleavage site (Henderson, et al. 1988) was used as a substrate (4E). The substrate in reaction buffer was mixed with aliquots of various cell extracts (see description of Figure 4 above) and incubated at $37^{\circ}C$. Equal eliquots of incubation mixture were taken at various time points and analyzed by RP-HPLC. The substrate in the 0 hour sample eluted as a single peak as shown in Figure 4A. After incubation for 1 hour, two newly appearing peaks, products labelled P1 and P2, can be seen, correlating with a significant decrease of the substrate peak. Subsequent amino acid analysis of the recovered peaks demonstrated that product 1 and product 2 corresponded to the expected cleavage products as shown in Table 1 proving a Tyr-Pro bond cleavage, which is the determined natural cleavage site. Extended incubation for 3 hours showed a further decrease of the substrate peak and substantial increase in the peak height of product 1,

indicating progression of the hydrolysis of the Tyr-Pro bond. However, the peak of product 1 seems to be smaller as expected since the absorbance of the tetrapeptide Pro-Ile-Val-Glu-NH_2 is substantially smaller than that of the pentapeptide having a free COOH-terminal tyrosine. An increase of product 1 and 2 after 3 hours of incubation showed a corresponding decrease of the substrate peak.

No cleavage products have been detected in reactions using extracts from uninduced cells, clone PR-C (Figure 4D) and of control cells (control plasmid PKK233-2; data not shown). There was no decrease in the substrate peak even after 6 hours of incubation (Figure 4D) indicating that the nonapeptide is resistent to degradation by bacterial proteases. This makes this substrate especially useful for assaying viral protease activities in crude extracts, facilitating purification and isolation of the protease.

The amino acid composition data for the substrate and its cleavage products are shown in Table 1. The amounts of observed amino acids correspond clearly to the expected amounts demonstrating that the cleavage occurs at the expected cleavage site of the synthetic peptide corresponding to the p17-p24 site of the gag precursor.

Table 1. Amino acid composition of the substrate and the cleavage products

Amino acids	Substrate		product 1		product 2	
	Predicted	Recovered	Predicted	Recovered	Predicted	Recovered
Asp	1	1.06	0	0.01	1	0.94
Glu	2	2.06	1	1.00	1	1.00
Ser	1	0.98	0	0.01	1	0.89
Pro	1	1.05	1	1.19	0	0.03
Tyr	1	1.06	0	0.01	1	1.01
Val	2	1.87	1	0.43*	1	1.16
Ile	1	0.92	1	0.45*	0	0.02

*The observed amounts of Val and Ile were found lower than expected in product 1 due to a frequently observed inefficient hydrolysis of the Ile-Val bond.

IN THE CLAIMS

1. A gene for encoding a protease of human immunodeficiency virus consisting essentially of:

a synthetic nucleotide sequence for a protease essential to infectivity of human immunodeficiency virus.

2. The gene of claim 1 wherein said protease is essential for infectivity of a retrovirus.

3. The gene of claim 2 wherein said retrovirus is a member of the group consisting of HIV-1, HIV-2, and HTLV a Human Leukemia virus.

4. A gene for encoding a protease of human immunodeficiency virus consisting essentially of:

a synthetic double stranded nucleotide sequence of which the coding sequence is:

10	20	30	40	50
CCTCAGATCA	CTCTTTGGCA	ACGACCCCTC	GTCACAATAA	AGATAGGGGG
60	70	80	90	100
GCAACTAAAG	GAAGCTCTAT	TAGATACAGG	AGCAGATGAT	ACAGTATTAG
110	120	130	140	150
AAGAAATGAG	TTTGCCAGGA	AGATGGAAAC	CAAAAATGAT	AGGGGGAATT
160	170	180	190	200
GGAGGTTTTA	TCAAAGTAAG	ACAGTATGAT	CAGATACTCA	TAGAAATCTG
210	220	230	240	250
TGGACATAAA	GCTATAGGTA	CAGTATTAGT	AGGACCTACA	CCTGTCAACA
260	270	280	290	
TAATTGGAAG	AAATCTGTTG	ACTCAGATTG	GTTGCACTTT	AAATTTT

5. A method for expressing a protease consisting essentially of inserting a recombinant vector containing a synthetic gene for a protease essential for infectivity of a retrovirus into a host cell;
expressing said gene; and
separating said protease.

6. The process of claim 5 wherein said retrovirus is a member of the group consisting of HIV-1, HIV-2, and HTLV a Human Leukemia virus.

7. The process of claim 5 wherein said retrovirus is HIV-1 and said gene has a nucleotide sequence of

10	20	30	40	50
CCTCAGATCA	CTCTTTGGCA	ACGACCCCTC	GTCACAATAA	AGATAGGGGG
60	70	80	90	100
GCAACTAAAG	GAAGCTCTAT	TAGATACAGG	AGCAGATGAT	ACAGTATTAG
110	120	130	140	150
AAGAAATGAG	TTTGCCAGGA	AGATGGAAAC	CAAAAATGAT	AGGGGGAATT
160	170	180	190	200
GGAGGTTTTA	TCAAAGTAAG	ACAGTATGAT	CAGATACTCA	TAGAAATCTG
210	220	230	240	250
TGGACATAAA	GCTATAGGTA	CAGTATTAGT	AGGACCTACA	CCTGTCAACA
260	270	280	290	
TAATTGGAAG	AAATCTGTTG	ACTCAGATTG	GTTGCACTTT	AAATTTT

FIG. 1

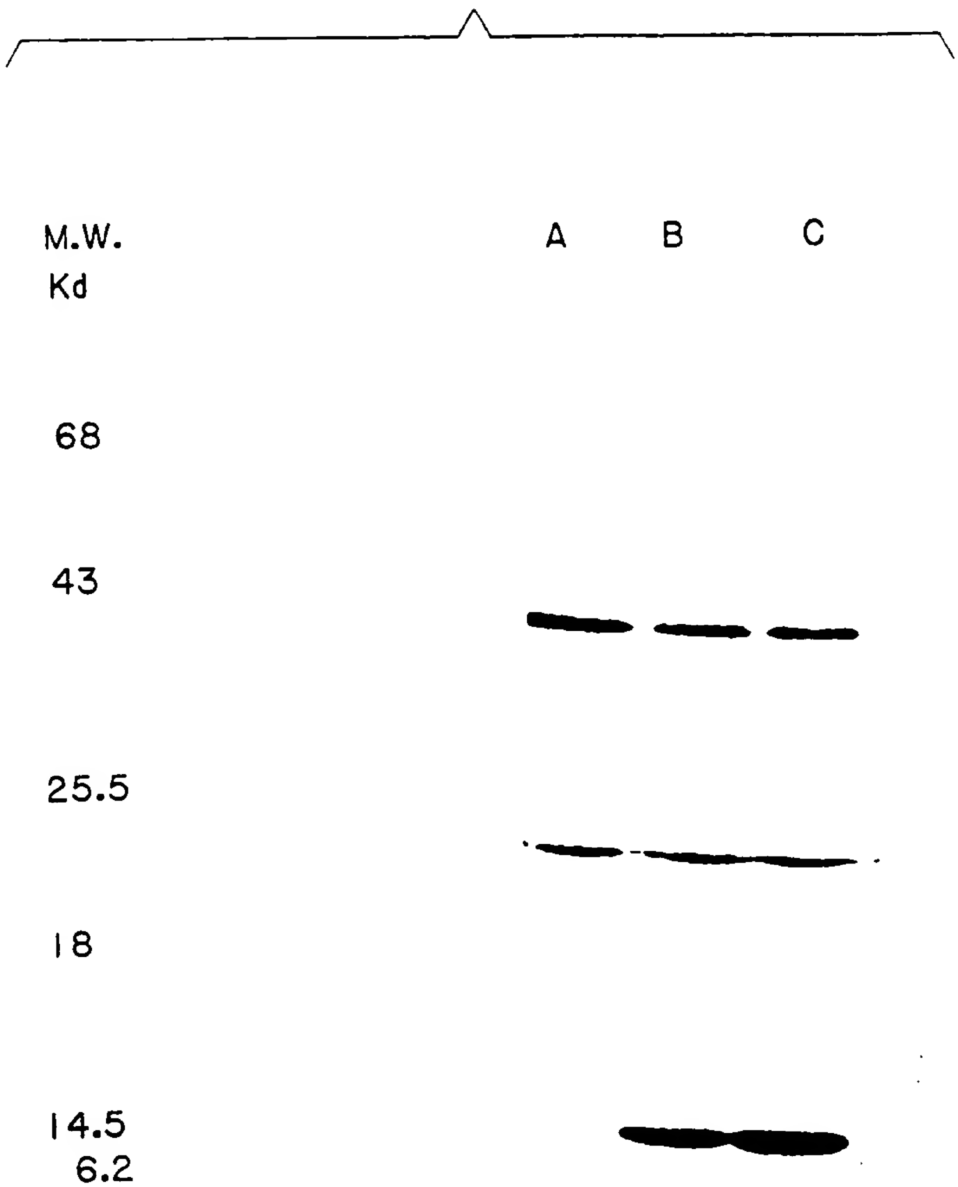


FIG. 2

fragment 1

5′ CCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGGGCAActaa

fragment 2

AGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAGTTTGCCAGGAagat

fragment 3

GGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAgata

fragment 4

CTCATAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTcaac

fragment 5

ATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTTAAATTTT 3′

FIG. 4E

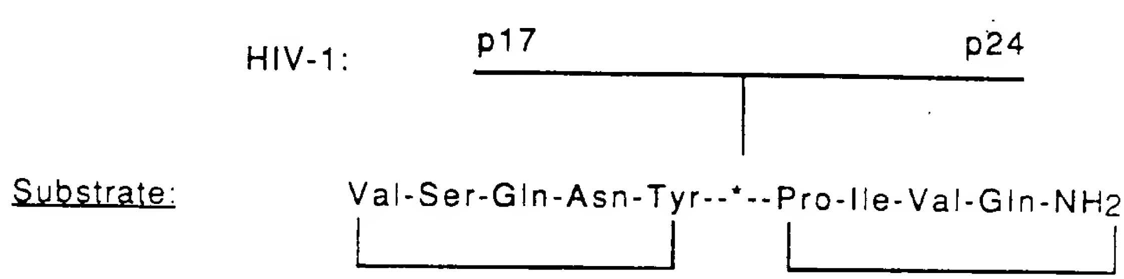


FIG. 3A FIG. 3B FIG. 3C

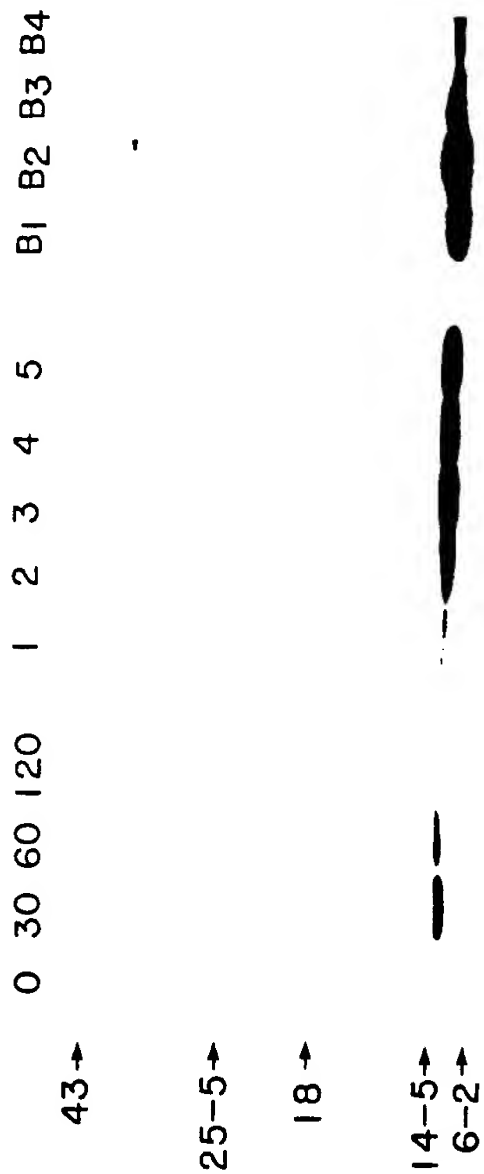
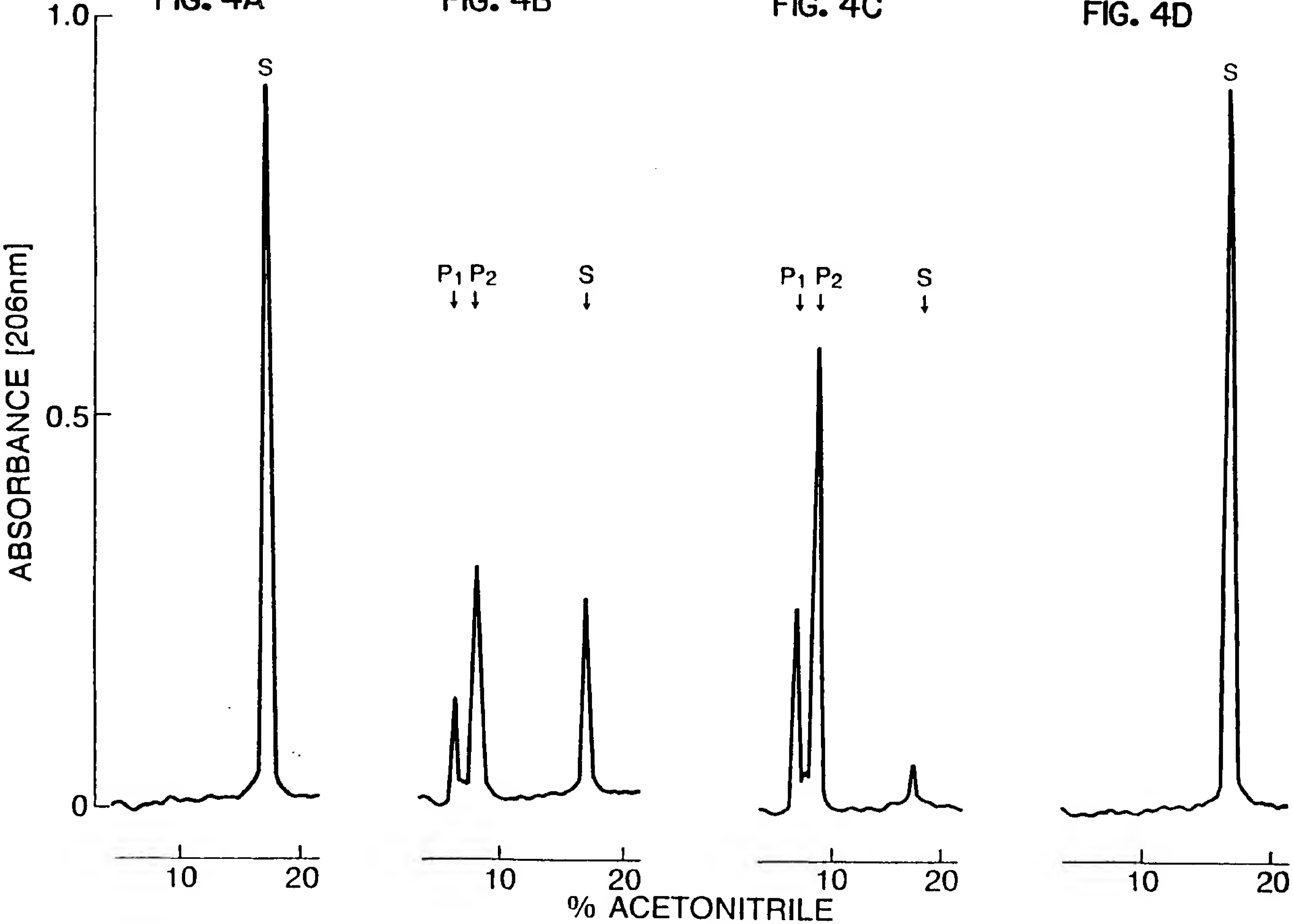


FIG. 4A FIG. 4B FIG. 4C FIG. 4D



INTERNATIONAL SEARCH REPORT

International Application No PCT/US89/02996

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(4): C 07 H 15/12; C12 N 9/48, 1/20, 15/00

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.	435/212,320,252.3; 536/27

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

CA File 1967-1989
BIOSIS File 1967-1989

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No [13]
X / Y	Nature, Volume 313, published 24 January, 1985. L. Ratner, et al. "Complete nucleotide sequence of the AIDS virus, HTLV-III," pp. 277-284. see section "second open reading frame", discussion section and Figures 1-4.	1-3 / 4-7
X / Y	Cell, Volume 40, published January 1985. S. Wain-Hobson, et al. "Nucleotide sequence of the AIDS virus, LAV," pp. 9-17. see section entitled "pol" and Figures 1-4	1-3 / 4-7
X / Y	Proc. Natl. Acad. Sci. USA, Volume 84, published December 1987. C. Debouk, et al. "Human immunodeficiency virus protease expressed in Escherichia Coli exhibits auto-processing and specific maturation of the gag precursor," pp. 8903-8906. see entire article	1-3 / 4-7
X / Y	Science, Volume 236, published 17 April 1987. W. Farmerie, et al. "Expression and processing of the AIDS virus reverse transcriptase in E. coli," pp. 305-308. see entire article	1-3 / 4-7

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
06 September 1989	28 NOV 1989

International Searching Authority	Signature of Authorized Officer
ISA/US	Beth A. Burrous

Form PCT/ISA/210 (second sheet) (Rev.11-87)